Estimation of a Panel Data Model with Parametric Temporal Variation in Individual Effects

Chirok Han
Victoria University of Wellington

Luis Orea
University of Oviedo

Peter Schmidt
Michigan State University

December 15, 2002

Abstract

This paper is an extension of Ahn, Lee and Schmidt (2001) to allow a *parametric* function for time-varying coefficients of the individual effects. It provides a fixed-effect treatment of models like those proposed by Kumbhakar (1990) and Battese and Coelli (1992). We present a number of GMM estimators based on different sets of assumptions. Least squares has unusual properties: its consistency requires white noise errors, and given white noise errors it is less efficient than a GMM estimator. We apply this model to the measurement of the cost efficiency of Spanish savings banks.

1 Introduction

In this paper we consider the model:

$$y_{it} = X_{it}'\beta + Z_i'\gamma + \lambda_t(\theta)\alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{1.1}$$

We treat T as fixed, so that "asymptotic" means as $N \to \infty$. The distinctive feature of the model is the interaction between the time-varying parametric function $\lambda_t(\theta)$ and the individual effect α_i. We consider the case that the α_i are "fixed effects," as will be discussed in more detail below. In this case estimation may be non-trivial due to the "incidental parameters problem" that the number of α's grows with sample size; see, for example, Chamberlain (1980).

Models of this form have been proposed and used in the literature on frontier productions functions (measurement of the efficiency of production). For example, Kumbhakar (1990) proposed the case that $\lambda_t(\theta) = [1 + \exp(\theta_1 t + \theta_2 t^2)]^{-1}$, and Battese and Coelli (1992) proposed the case that $\lambda_t(\theta) = \exp\big(-\theta(t - T)\big)$. Both of these papers considered random effects models in which α_i is independent of X and Z. In fact, both of these papers proposed specific (truncated normal) distributions for the α_i, with estimation by maximum likelihood. The aim of the present paper is to provide a fixed-effects treatment of models of this type.

There is also a literature on the case that the λ_t themselves are treated as parameters. That is, the model becomes:

$$y_{it} = X_{it}'\beta + Z_i'\gamma + \lambda_t\alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{1.2}$$

This corresponds to using a set of dummy variables for time rather than a parametric function $\lambda_t(\theta)$, and now $\lambda_t\alpha_i$ is just the product of fixed time and individual effects. This model has been considered by Kiefer (1980), Holtz-Eakin, Newey and Rosen (1988), Lee (1991), Chamberlain (1992), Lee and Schmidt (1993) and Ahn, Lee and Schmidt (2001), among others. Lee (1991) and Lee and Schmidt (1993) have applied this model to the frontier production function problem, in order to avoid having to assume a specific parametric function $\lambda_t(\theta)$. Another motivation for the model is that a fixed-effects version allows one to control for unobservables (e.g. macro events) that are the same for each individual, but to which different individuals may react differently.

Ahn, Lee and Schmidt (2001) establish some interesting results for the estimation of model (1.2). A generalized method of moments (GMM) estimator of the type considered by Holtz-Eakin, Newey and Rosen (1988) is consistent given exogeneity assumptions on the regressors X and Z. Least squares applied to (1.2), treating the α_i as fixed parameters, is consistent provided that the regressors are strictly exogenous and that the errors ϵ_{it} are white noise. The requirement of white noise errors for consistency of least squares is unusual, and is a reflection of the incidental parameters problem. Furthermore, if the errors are white noise, then a GMM estimator that incorporates the white noise assumption dominates least squares, in the sense of being asymptotically more efficient. This is also a somewhat unusual result, since in the usual linear model with normal errors, the moment conditions implied by the white noise assumption would not add to the efficiency of estimation.

The results of Ahn, Lee and Schmidt apply only to the case that the λ_t are unrestricted, and therefore do not apply to the model (1.1). However, in this paper we show that essentially the same results do hold for the model (1.1). This enables us to use a parametric function $\lambda_t(\theta)$, and to test the validity of this assumption, while maintaining only weak assumptions on the α_i. This may be very useful, especially in the frontier production function setting. Applications using unrestricted λ_t have yielded temporal patterns of efficiency that seem unreasonably variable and in need of smoothing, which a parametric function can accomplish.

The plan of the paper is as follows. Section 2 restates the model and lists our assumptions. Section 3 considers GMM estimation under basic exogeneity assumptions, while Section 4 considers GMM when we add the conditions implied by white noise errors. Section 5 considers least squares estimation and the sense in which it is dominated by GMM. In Section 6, this methodology is applied to the measurement of cost efficiency of Spanish banks. Finally, Section 7 contains some concluding remarks.

2 The Model and Assumptions

The model is given in equation (1.1) above. We can rewrite it in matrix form, as follows. Let $y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (X_{i1}, \ldots, X_{iT})'$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$. Thus y_i is $T \times 1$, X_i is $T \times K$, ϵ_i is $T \times 1$, β is $K \times 1$, γ is $g \times 1$, and α_i is a scalar. (In this paper, all the vectors are column vectors, and the data matrices are "vertically tall.") Define a function $\lambda : \Theta \to \mathbb{R}^T$, where Θ is a compact subset of $\mathbb{R}^p$, such that $\lambda(\theta) = (\lambda_1(\theta), \ldots, \lambda_T(\theta))'$. Note that T is fixed.

In matrix form, our model is:

$$y_i = X_i\beta + 1_T Z_i'\gamma + \lambda(\theta)\alpha_i + \epsilon_i, \quad i = 1, \ldots, N. \tag{2.1}$$

$\lambda(\theta)$ must be normalized in some way such as $\lambda(\theta)'\lambda(\theta) \equiv 1$ or $\lambda_1(\theta) \equiv 1$, to rule out trivial failure of identification arising from $\lambda(\theta) = 0$ or scalar multiplications of $\lambda(\theta)$. Here we choose the normalization $\lambda_1(\theta) \equiv 1$.

We assume that $p \leq T - 1$. When $p < T - 1$, our parametric specification for $\lambda(\theta)$ restricts the temporal pattern of λ. However, this model also includes the model of Ahn, Lee and Schmidt (2001) as the special case corresponding to $p = T - 1$ and $\lambda_t(\theta) = \theta_t$, $t = 2, \ldots, T$ (with $\lambda_1(\theta) = 1$ as above).

Let $W_i = (X_{i1}', \ldots, X_{iT}', Z_i')'$. We make the following "orthogonality" and "covariance" assumptions.

Assumption 1 (Orthogonality). $E(W_i', \alpha_i)'\epsilon_i' = 0$.

Assumption 2 (Covariance). $E\epsilon_i\epsilon_i' = \sigma_\epsilon^2 I_T$.

Assumption 1 says that ϵ_{it} is uncorrelated with α_i, Z_i, and $X_{i1}, \ldots, X_{iT}$, and therefore contains an assumption of strict exogeneity of the regressors. Note that it does not restrict the correlation between α_i and $[Z_i, X_{i1}, \ldots, X_{iT}]$, so that we are in the fixed-effects framework. Assumption 2 asserts that the errors are white noise.

We also assume the following regularity conditions.

Assumption 3 (Regularity).

(*i*) $(W_i', \alpha_i, \epsilon_i')'$ *is independently and identically distributed over* i*;*

(*ii*) ϵ_i *has finite fourth moment, and* $E\epsilon_i = 0$*;*

(*iii*) $(W_i', \alpha_i)'$ *has finite nonsingular second moment matrix;*

(*iv*) $EW_i(Z_i', \alpha_i)$ *is of full column rank;*

(*v*) $\lambda(\theta)$ *is twice continuously differentiable in* θ.

The first four of these conditions correspond to assumptions (BA.1)–(BA.4) of Ahn, Lee and Schmidt (2001), who give some explanation. We note that condition (*iii*) requires that the effect α_i be correlated with some variable in W_i. This condition is needed for identification under the Orthogonality Assumption only. Condition (*v*) is new, and self-explanatory.

3 GMM under the Orthogonality Assumption

Let $u_{it} = u_{it}(\beta, \gamma) = y_{it} - X_{it}'\beta - Z_i'\gamma$, and $u_i = u_i(\beta, \gamma) = (u_{i1}, \ldots, u_{iT})'$. Since $u_{it} = \lambda_t(\theta)\alpha_i + \epsilon_{it}$, it follows that $u_{it} - \lambda_t(\theta)u_{i1} = \epsilon_{it} - \lambda_t(\theta)\epsilon_{i1}$, which does not depend on α_i. This is a sort of generalized within transformation to remove the individual effects. The Orthogonality Assumption (Assumption 1) then implies the following moment conditions:

$$EW_i[u_{it}(\beta, \gamma) - \lambda_t(\theta)u_{i1}(\beta, \gamma)] = 0, \; t = 2, \ldots, T. \tag{3.1}$$

These moment conditions can be written in matrix form, as follows. Define $G(\theta) = [-\lambda_*(\theta), I_{T-1}]'$, where $\lambda_* = (\lambda_2, \ldots, \lambda_T)'$. The generalized within transformation corresponds to multiplication by $G(\theta)'$, and the moment conditions (3.1) can equivalently be written as follows:

$$Eb_{1i}(\beta, \gamma, \theta) = E[G(\theta)' u_i(\beta, \gamma) \otimes W_i] = 0. \tag{3.2}$$

(This corresponds to equation (7) of Ahn, Lee and Schmidt (2001), but looks slightly different because our W_i is a column vector whereas theirs is a row vector.) This is a set of $(T-1)(TK+g)$ moment conditions.

Some further analysis is needed to establish that (3.2) contains *all* of the moment conditions implied by the Orthogonality Assumption. Let $\Sigma_{WW} = EW_iW_i'$, $\Sigma_{W\alpha} = EW_i\alpha_i$, and $\sigma_\alpha^2 = E\alpha_i^2$. Given the model (2.1), the Orthogonality Assumption holds if and only if the following moment conditions hold:

$$E[u_i(\beta, \gamma) \otimes W_i - \lambda(\theta) \otimes \Sigma_{W\alpha}] = 0. \tag{3.3}$$

We could use these moment conditions as the basis for GMM estimation. Alternatively, we can remove the parameter $\Sigma_{W\alpha}$ by applying a nonsingular linear transformation to (3.3) in such a way that the transformed set of moment conditions is separated into two subsets, where the first subset does not contain $\Sigma_{W\alpha}$ and the second subset is exactly identified for $\Sigma_{W\alpha}$, given (β, γ, θ). The following transformation accomplishes this.

$$E \begin{bmatrix} G' \otimes I_d \\ \lambda' \otimes I_d \end{bmatrix} [u_i \otimes W_i - \lambda \otimes \Sigma_{W\alpha}] = 0 \tag{3.4}$$

where $d \equiv TK + g$ for notational simplicity; similarly, G, λ and u_i are shortened expressions for $G(\theta)$, $\lambda(\theta)$ and $u_i(\beta, \gamma)$. This is a nonsingular transformation, since (G, λ) is nonsingular, and therefore GMM based on (3.4) is asymptotically equivalent to GMM based on (3.3). Now split

(3.4) into its two parts:

$$E(G'u_i \otimes W_i) = 0 \tag{3.5}$$

$$E(\lambda' u_i)W_i - (\lambda'\lambda)\Sigma_{W\alpha} = 0. \tag{3.6}$$

Here (3.6) is exactly identified for $\Sigma_{W\alpha}$, given β, γ and θ, in the sense that the number of moment conditions in (3.6) is the same as the dimension of $\Sigma_{W\alpha}$. Also $\Sigma_{W\alpha}$ does not appear in (3.5). It follows (e.g., Ahn and Schmidt (1995), Theorem 1) that the GMM estimates of β, γ and θ from (3.5) alone are the same as the GMM estimates of β, γ and θ if we use both (3.5) and (3.6), and estimate the full set of parameters $(\beta, \gamma, \theta, \Sigma_{W\alpha})$. But (3.5) is the same as (3.2), which establishes that (3.2) contains all the useful information about β, γ and θ implied by the Orthogonality Assumption.

Let $\bar{b}_1(\beta, \gamma, \theta) = N^{-1} \sum_{i=1}^{N} b_{1i}(\beta, \gamma, \theta)$. Then the optimal GMM estimator $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\theta}$ based on the Orthogonality Assumption solves the problem

$$\min_{\beta,\gamma,\theta} N\bar{b}_1(\beta, \gamma, \theta)' V_{11}^{-1} \bar{b}_1(\beta, \gamma, \theta) \tag{3.7}$$

where $V_{11} = Eb_{1i}b_{1i}'$ evaluated at the true parameters. As usual, V_{11} can be replaced by any consistent estimate. A standard estimate would be

$$\hat{V}_{11} = \frac{1}{N} \sum_{i=1}^{N} b_{1i}(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta}) b_{1i}(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta})' \tag{3.8}$$

where $(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta})$ is an initial consistent estimate of (β, γ, θ) such as GMM using identity weighting matrix. Under certain regularity conditions (Hansen (1982), Assumption 3) the resulting GMM estimator is $\sqrt{N}$-consistent and asymptotically normal.

To express the asymptotic variance of the GMM estimator analytically, we need a little more notation. Let S_X be the $T(TK+g) \times K$ selection matrix such that $X_i = (I_T \otimes W_i)' S_X$, and let S_Z be the $T(TK+g) \times g$ selection matrix such that $1_T Z_i' = (I_T \otimes W_i)' S_Z$. S_X and S_Z have the following forms:

$$S_X = (I_K \; O \; \cdots \; O \; O_{K\times g} \vdots O \; I_K \; \cdots \; O \; O_{K\times g} \vdots \cdots \vdots O \; O \; \cdots \; I_K \; O_{K\times g})' \tag{3.9a}$$

$$S_Z = (O_{g\times K} \; \cdots \; O_{g\times K} \; I_g \vdots \cdots \vdots O_{g\times K} \; \cdots \; O_{g\times K} \; I_g)' = 1_T \otimes (O_{g\times TK}, I_g)' \tag{3.9b}$$

where O's without dimension subscript stand for $O_{K\times K}$. Define $\Lambda_* = \partial\lambda_*(\theta_0)/\partial\theta'$. The variance of the asymptotic distribution of the GMM estimates of β, γ and θ equals $(B_1' V_{11}^{-1} B_1)^{-1}$ where $V_{11} = Eb_{1i}b_{1i}'$ as above and

$$B_1 = -[(G \otimes \Sigma_{WW})' S_X, (G \otimes \Sigma_{WW})' S_Z, \Lambda_* \otimes \Sigma_{W\alpha}]. \tag{3.10}$$

This result can be obtained either by direct calculation, or by applying the chain rule to B_1 calculated in Ahn, Lee and Schmidt (2001, p. 251). This asymptotic variance form is obtained from the Orthogonality Assumption only and does not need any further assumption.

A consistent estimate of B_1 can be obtained as

$$\hat{B}_1 = \frac{1}{N}\sum_{i=1}^{N} \hat{B}_{1i}, \quad \hat{B}_{1i} = -(\hat{G}' X_i, \hat{G}' 1_T Z_i', \hat{u}_{i1}\hat{\Lambda}_*) \otimes W_i. \tag{3.11}$$

Here $\hat{B}_{1i}$ is the matrix of first derivatives of b_{1i} with respect to the parameters (β, γ, θ), evaluated at the GMM estimates.

A practical problem with this GMM procedure is that it is based on a rather large set of moment conditions. For example, in our empirical analysis, b_{1i} will reflect 846 moment conditions. One might want to reduce this number by considering only a subset of the moment conditions. One possibility is to replace the instruments W_i in (3.2) by P_i, where P_i is a subset of W_i. This possibility is discussed in Appendix A.

Alternatively, we can reduce the number of moment conditions considerably without sacrificing efficiency of estimation if we make the following assumption of no conditional heteroskedasticity (NCH) of ϵ_i:

$$E(\epsilon_i \epsilon_i' | W_i) = \Sigma_{\epsilon\epsilon}. \tag{NCH}$$

Under the NCH assumption,

$$V_{11} = E[G(\theta_0)'\epsilon_i\epsilon_i'G(\theta_0) \otimes W_iW_i'] = G(\theta_0)'\Sigma_{\epsilon\epsilon}G(\theta_0) \otimes \Sigma_{WW}. \tag{3.12}$$

Using this result, the set of moment conditions (3.2) can be converted into an *exactly identified* set of moment conditions that yield an asymptotically equivalent GMM estimate. Specifically, we can replace the moment conditions $Eb_{1i} = 0$ by the moment conditions $EB_1'V_{11}^{-1}b_{1i} = 0$. Routine calculation using the forms of B_1, V_{11} and b_{1i} yields the explicit expression:

$$EX_i'G(G'\Sigma_{\epsilon\epsilon}G)^{-1}G'u_i = 0 \tag{3.13a}$$

$$EZ_i1_T'G(G'\Sigma_{\epsilon\epsilon}G)^{-1}G'u_i = 0 \tag{3.13b}$$

$$E\Sigma_{W\alpha}'\Sigma_{WW}^{-1}W_i \cdot \Lambda_*'(G'\Sigma_{\epsilon\epsilon}G)^{-1}G'u_i = 0. \tag{3.13c}$$

These three sets of moment conditions respectively correspond to (21a), (21b), and (21c) of Ahn, Lee and Schmidt (2001, p. 229). The point of this simplification is that we have drastically reduced the set of moment conditions: there are $(T-1)(TK+g)$ moment conditions in b_{1i} (equation (3.2)) but only $K + g + p$ moment conditions in (3.13).

We note that this is a stronger result than the corresponding result (Proposition 1, p. 229) of Ahn, Lee and Schmidt (2001). In order to reach essentially the same conclusion on the reduction of the number of moment conditions, they impose the assumption that ϵ_i is independent of (W_i, α_i), a much stronger assumption than our NCH assumption.

In order to make this procedure operational, we need to replace the nuisance parameters $\Sigma_{\epsilon\epsilon}$, $\Sigma_{W\alpha}$ and Σ_{WW} by consistent estimates, based on some initial consistent GMM estimates of β, γ and θ. Σ_{WW} can be consistently estimated by $\hat{\Sigma}_{WW} = N^{-1}\sum_{i=1}^{N} W_iW_i'$. Also, for any sequence (β_N, γ_N) that converges in probability to (β_0, γ_0), we have

$$\frac{1}{N}\sum_{i=1}^{N} u_i(\beta_N, \gamma_N)u_i(\beta_N, \gamma_N)' \xrightarrow{p} \Sigma_{\epsilon\epsilon} + \sigma_\alpha^2\lambda(\theta_0)\lambda(\theta_0)'. \tag{3.14}$$

Since $G(\theta)'\lambda(\theta) = 0$, for any initial consistent estimate $(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta})$,

$$G(\tilde{\theta})' \left(N^{-1} \sum_{i=1}^{N} u_i(\tilde{\beta}, \tilde{\gamma}) u_i(\tilde{\beta}, \tilde{\gamma})' \right) G(\tilde{\theta}) \tag{3.15}$$

will consistently estimate $G(\theta_0)'\Sigma_{\epsilon\epsilon}G(\theta_0)$. Thus it is easy to construct a consistent estimate of V_{11} as given in (3.12).

In order to consistently estimate the asymptotic variance under NCH, we need to estimate Σ_{WW}, $\Sigma_{W\alpha}$, and $G'\Sigma_{\epsilon\epsilon}G$. Estimation of Σ_{WW} and $G'\Sigma_{\epsilon\epsilon}G$ was discussed above. We can obtain an estimate of $\Sigma_{W\alpha}$ from the GMM problem (3.4). A direct algebraic calculation gives us that

$$\hat{\Sigma}_{W\alpha} = \frac{1}{N} \sum_{i=1}^{N} W_i \frac{\hat{\lambda}'\hat{u}_i}{\hat{\lambda}'\hat{\lambda}} - \frac{1}{N} \sum_{i=1}^{N} W_i [\widehat{\lambda'\Sigma_{\epsilon\epsilon}G} (\widehat{G'\Sigma_{\epsilon\epsilon}G})^{-1} \hat{G}'\hat{u}_i]/(\hat{\lambda}'\hat{\lambda}) \tag{3.16}$$

where $\hat{u}_i = u_i(\hat{\beta}, \hat{\gamma})$, $\hat{\lambda} = \lambda(\hat{\theta})$, $\hat{G} = G(\hat{\theta})$, and $\widehat{\lambda'\Sigma_{\epsilon\epsilon}G}$ is a consistent estimate of $\lambda'\Sigma_{\epsilon\epsilon}G$, one possibility of which is $N^{-1}\sum_{i=1}^{N} \hat{\lambda}'\hat{u}_i\hat{u}_i'\hat{G}$.

It is important to observe that the moment conditions (3.13) are linear combinations of the moment conditions b_{1i} in (3.2), and therefore they are valid moment conditions under the Orthogonality Assumption only. That is, these moment conditions hold and can be used as a valid basis of GMM estimation so long as the Orthogonality Assumption holds, whether or not the NCH assumption holds. The set of moment conditions in (3.2) may be very large, and so the simplification involved in using the exactly identified (minimal size) set of moment conditions (3.13) may be useful in practice. The only point of the NCH assumption is that, if it holds, the conditions (3.13) are the *optimal* exactly identified set of moment conditions, so that GMM using (3.13) is just as efficient as GMM using the full set of moment conditions given in (3.2). If the NCH condition does not hold, we can still base GMM on (3.13), but there is a loss of efficiency relative to using the full set (3.2).

We have already discussed how to estimate the variance matrix of the GMM estimator under the NCH assumption. However, because the moment conditions (3.13) are still valid without the

NCH assumption, it is useful to have an estimate of the variance of the GMM estimate that is consistent whether or not the NCH condition holds. Standard methods starting with the moment conditions (3.13) should yield such an estimate. Some details are given in Appendix B.

4 GMM under the Orthogonality and Covariance Assumptions

In this section we continue to maintain the Orthogonality Assumption (Assumption 1), but now we add the Covariance Assumption (Assumption 2), which asserts that $E\epsilon_i\epsilon_i' = \sigma_\epsilon^2 I_T$.

Clearly the Covariance Assumption holds if and only if

$$E(u_i u_i') = \sigma_\alpha^2 \lambda\lambda' + \sigma_\epsilon^2 I_T. \tag{4.1}$$

Condition (4.1) contains $T(T+1)/2$ distinct moment conditions. It also contains the two nuisance parameters σ_α^2 and σ_ϵ^2, and so it should imply $T(T+1)/2-2$ moment conditions for the estimation of β, γ and θ. These are in addition to the moment conditions (3.2) implied by the Orthogonality Assumption.

To write these moment conditions explicitly, we need to define some notation. Let $H = \mathrm{diag}(H_2, H_3, \ldots, H_T)$, with H_t equal to the $T \times (T-t)$ matrix of the last $T-t$ columns (the $(t+1)$th through Tth columns) of I_T for $t < T$, and with H_T equal to a $T \times (T-2)$ matrix of the second through $(T-1)$-th columns of I_T.[1] Then we can write the distinct moment conditions implied by the Orthogonality and Covariance Assumptions as follows:

$$Eb_{1i} = E(G'u_i \otimes W_i) = 0 \tag{4.2a}$$

$$Eb_{2i} = EH'(G'u_i \otimes u_i) = 0 \tag{4.2b}$$

$$Eb_{3i} = E\big[G'u_i \otimes \frac{\lambda' u_i}{\lambda'\lambda}\big] = 0. \tag{4.2c}$$

[1] For any matrix B with T rows, $H_t'B$ selects the last $T-t$ rows of B for $t < T$, and $H_T'B$ selects the second through $(T-1)$-th rows of B. For any matrix B with T columns, BH_t selects the last $T-t$ columns of B for $t < T$, and BH_T selects the second through $(T-1)$-th columns of B.

(In these expressions, G is short for $G(\theta)$, λ is short for $\lambda(\theta)$, and u_i is short for $u_i(\beta, \gamma)$.)

The moment conditions b_{1i} in (4.2a) are exactly the same as those in (3.2) of the previous section, and follow from the Orthogonality Assumption.

The moment conditions b_{2i} in (4.2b) correspond to those in equation (12) of Ahn, Lee and Schmidt (2001). Note that it is not the case that $E(G'u_i \otimes u_i) = 0$. Rather, looking at a typical element of this product, we have $E(u_{it} - \lambda_t u_{i1})u_{is}$, which equals zero for $s \neq t$ and $s \neq 1$. The selection matrix H' picks out the logically distinct products of expectation zero, the number of which equals $T(T-1)/2 - 1$. The selection matrix H plays the same role as the definition of the matrices U_{it}° plays in Ahn, Lee and Schmidt (2001). We note that the moment conditions b_{2i} follow from the non-autocorrelation of the ϵ_{it}; homoskedasticity would not be needed.

The $(T-1)$ moment conditions in b_{3i} in (4.2c) correspond to those in equation (13) of Ahn, Lee and Schmidt (2001). They assert that, for $t = 2, \ldots, T$, $E(u_{it} - \lambda_t u_{i1})(\sum_{s=1}^{T} \lambda_s u_{is}) = 0$, and their validity depends on both the non-autocorrelation and the homoskedasticity of the ϵ_{it}.

Some further analysis may be useful to establish that (4.2b) and (4.2c) represent all of the useful implications of the Covariance Assumption. We begin with the implication (4.1) of the Covariance Assumption, which we rewrite as

$$E(u_i \otimes u_i) = \sigma_\alpha^2(\lambda \otimes \lambda) + \sigma_\epsilon^2 \text{vec} I_T. \tag{4.3}$$

Now, let S be the $T^2 \times T(T+1)/2$ selection matrix such that, for a $T \times 1$ vector u, $\text{vech}(uu') = S'(u \otimes u)$, where "vech" is the vector of distinct elements. Then

$$ES'(u \otimes u) = S'[\sigma_\alpha^2(\lambda \otimes \lambda) + \sigma_\epsilon^2 \text{vec} I_T] \tag{4.4}$$

contains the distinct moment conditions.

Now we transform the moment conditions (4.4) by multiplying them by a nonsingular matrix, in such a way that (i) the first $T(T+1)/2 - 2$ transformed moment conditions are those given in

(4.2b) and (4.2c); and (ii) the last two moment conditions are exactly identified for the nuisance parameters (σ_α^2 and σ_ϵ^2), given the other parameters. This will imply that the last two moment conditions are redundant for the estimation of β, γ and θ, and thus that (4.2b) and (4.2c) contain all of the useful information implied by the Covariance Assumption for estimation of β, γ and θ.

To exhibit the transformation, let G_t be the $(t-1)$th column of G; let e_t^* equal the tth column of I_{T-2} and e_T equal the last column of I_T; and define

$$(H_T^{**})' = [-\lambda_T H_T',\ e_1^* e_T',\ \ldots,\ e_{T-2}^* e_T',\ O_{(T-2)\times T}]. \tag{4.5}$$

(H_T was defined above.) Then

$$[G_2 \otimes H_2,\ \ldots,\ G_{T-1} \otimes H_{T-1},\ H_T^{**}]' S \cdot S'(u_i \otimes u_i) = H'(G' \otimes I_T)(u_i \otimes u_i), \tag{4.6}$$

which is the same as in b_{2i} in (4.2b). Also, let $J_1^* = I_T - \lambda\lambda'$ and J_t^*, $t = 2, \ldots, T$, is equal to $\mathrm{diag}\{O_{t\times t}, \lambda_t I_{T-t}\}$ plus a $T \times T$ matrix with zero elements except for the tth row which is λ'. Then

$$H_1'[J_1^*, \ldots, J_T^*] S \cdot S'(u_i \otimes u_i) = (\lambda' \otimes G')(u_i \otimes u_i), \tag{4.7}$$

which is equal to b_{3i} in (4.2c).

The point of the above argument is that the transformations preceding $S'(u_i \otimes u_i)$ in (4.6) and (4.7), stacked vertically, construct a $[T(T+1)/2 - 2] \times T(T+1)/2$ matrix of full row rank, and yield the moment conditions b_{2i} and b_{3i}. The remaining two moment conditions that determine the nuisance parameters are

$$E \begin{bmatrix} u_{i1}^2 \\ u_{i2}u_{i1} \end{bmatrix} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 \\ \lambda_2 \sigma_\alpha^2 \end{bmatrix} \tag{4.8}$$

and must be linearly independent of the others (since they involve σ_α^2 and σ_ϵ^2 while the others do not).

The set of moment conditions (4.2) may be large, since the number of moment conditions in (4.2a) may be large. As in the previous section, we can reduce this number by using only a subset P_i of the instruments W_i. This is discussed in Appendix A. Alternatively, we can simplify things with the following "conditional independence of the moments up to fourth order" (CIM4) assumption:

> Conditional on (W_i, α_i), ϵ_{it} is independent over $t = 1, 2, \ldots, T$, with mean zero, and with second, third and fourth moments that do not depend on (W_i, α_i) or on t. (CIM4)

This is a strong assumption; it implies the Orthogonality Assumption, the Covariance Assumption, the NCH assumption, and more. In Appendix C, we calculate the asymptotic variance matrix of the GMM estimate based on (4.2) under the assumption (CIM4). More fundamentally, if assumption (CIM4) is true, we can reduce the number of moment conditions without reducing efficiency of estimation. Specifically, let $\Lambda = \partial\lambda(\theta_0)/\partial\theta$ and note that $\Lambda_* = G'\Lambda$. Given assumption (CIM4), the moment conditions (3.13), which are asymptotically equivalent to (4.2a), can be simplified as follows:

$$EX_i'P_G u_i = 0 \tag{4.9a}$$

$$EZ_i 1_T' P_G u_i = 0 \tag{4.9b}$$

$$E\Sigma_{W\alpha}'\Sigma_{WW}^{-1}W_i \cdot \Lambda' P_G u_i = 0. \tag{4.9c}$$

That is, in place of the large set of moment conditions (4.2a), (4.2b) and (4.2c), we can use the reduced set of moment conditions consisting of (4.9), (4.2b) and (4.2c).

We can note that, when $\Sigma_{\epsilon\epsilon} = \sigma_\epsilon^2 I$, the moment conditions (3.13) are the same as (4.9). This is not surprising since, if the CIM4 assumption is true, so is the NCH assumption.

A final simplification arises if, conditional on (W_i, α_i), ϵ_{it} is i.i.d. normal. In this case, (4.2b) can be shown to be redundant given (4.2a) and (4.2c). (See Proposition 4 of Ahn, Lee and Schmidt (2001, p. 231).) Hence, in that case, the GMM estimator using the moment conditions (4.9) and (4.2c) is efficient.

We note that the simplifications that arise here, given the CIM4 assumption or the i.i.d. normal assumption, are similar in spirit to those that arose in Section 3 under the NCH assumption. For example, the set of moment conditions consisting of (4.9), (4.2b) and (4.2c) is much smaller than the full set (4.2), and this simplified set of moment conditions may be useful in practice, whether or not the CIM4 assumption holds. The point of the CIM4 assumption is simply that it identifies the circumstances under which we can use the reduced set of moment conditions without a loss of efficiency. If the CIM4 assumption does not hold, the GMM estimator using the reduced set of moment conditions is still consistent (so long as the Orthogonality and Covariance Assumptions hold), but it would be less efficient than the GMM estimator using the full set of moment conditions (4.2). Similar comments apply to the simplification that arises from dropping (4.2b): we can always do this, but it causes a loss of efficiency if the i.i.d. normal assumption does not hold.

In Appendix B we show how to calculate an estimate of the variance of these GMM estimates that is consistent whether or not the CIM4 or i.i.d. normal assumptions hold.

5 Least Squares

In this section we consider the concentrated least squares (CLS) estimation of the model. We treat the α_i as parameters to be estimated, so this is a true "fixed effects" treatment. We can consider the following least squares problem:

$$\min_{\beta,\gamma,\theta,\alpha_1,\ldots,\alpha_N} N^{-1}\sum_{i=1}^{N}[y_i - X_i\beta - 1_T Z_i'\gamma - \lambda(\theta)\alpha_i]'[y_i - X_i\beta - 1_T Z_i'\gamma - \lambda(\theta)\alpha_i]. \tag{5.1}$$

Solving for $\alpha_1, \ldots, \alpha_N$ first, we get

$$\alpha_i(\beta,\gamma,\theta) = [\lambda(\theta)'\lambda(\theta)]^{-1}\lambda(\theta)'u_i(\beta,\gamma) \quad i = 1, \ldots, N. \tag{5.2}$$

where $u_i(\beta,\gamma) = y_i - X_i\beta - 1_T Z_i'\gamma$ as before. Then the estimates $\hat{\beta}_{LS}$, $\hat{\gamma}_{LS}$, and $\hat{\theta}_{LS}$ minimizing (5.1) are equal to the minimizers of the *sum of the squared concentrated residuals*

$$\bar{C}(\beta,\gamma,\theta) = N^{-1}\sum_{i=1}^{N} C_i(\beta,\gamma,\theta) = N^{-1}\sum_{i=1}^{N} u_i(\beta,\gamma)'M_{\lambda(\theta)}u_i(\beta,\gamma) \tag{5.3}$$

which is obtained by replacing α_i in (5.1) with (5.2). From the name of (5.3), we call $\hat{\beta}_{LS}$, $\hat{\gamma}_{LS}$ and $\hat{\theta}_{LS}$ the *concentrated least squares estimator*.

Since $G'\lambda = 0$, we have $M_\lambda G = G$ and therefore $M_\lambda = P_G = G(G'G)^{-1}G'$. So the first order conditions of the CLS estimation become

$$\begin{bmatrix} \partial\bar{C}/\partial\beta \\ \partial\bar{C}/\partial\gamma \\ \partial\bar{C}/\partial\theta \end{bmatrix} = -\frac{2}{N}\sum_{i=1}^{N}\begin{bmatrix} X_i'P_G u_i \\ Z_i 1_T' P_G u_i \\ \Lambda' P_G u_i u_i'\lambda(\lambda'\lambda)^{-1} \end{bmatrix} = 0. \tag{5.4}$$

Interpreting (5.4) as sample moment conditions, we can construct the corresponding (exactly identified) implicit population moment conditions:

$$EX_i'P_G u_i = 0 \tag{5.5a}$$

$$EZ_i 1_T' P_G u_i = 0 \tag{5.5b}$$

$$E\Lambda' P_G u_i u_i'\lambda(\lambda'\lambda)^{-1} = 0. \tag{5.5c}$$

That is, the CLS estimator is asymptotically equivalent to the GMM estimator based on (5.5).

The moment conditions (5.5a) and (5.5b) are satisfied under the Orthogonality Assumption. However, this is not true of (5.5c). The moment conditions (5.5c) require the Covariance Assumption to be valid (unless we make very specific and unusual assumptions about the form of λ and its relationship to the error variance matrix). Thus, the consistency of the CLS estimator requires *both* the Orthogonality Assumption *and* the Covariance Assumption. This is a rather striking result, since the consistency of least squares does not usually require restrictions on the second moments of the errors, and is a reflection of the incidental parameters problem.

We would generally believe that least squares should be efficient when the errors are i.i.d. normal. However, similarly to the result in Ahn, Lee and Schmidt (2001), this is not true in the present case. The efficient GMM estimator under the Orthogonality and Covariance Assumptions uses the moment conditions (4.2), while the CLS estimator uses only a subset of these. This can be seen most explicitly in the case that, conditional on (W_i, α_i), the ϵ_{it} are i.i.d. normal. Then (4.2b) is redundant and (4.2a) can be replaced by (4.9), so that the efficient GMM estimator is based on (4.9a), (4.9b), (4.9c) and (4.2c). The CLS estimator is based on (5.5a), which is the same as (4.9a); (5.5b), which is the same as (4.9b); and (5.5c), which is a subset of (4.2c).[2] So the inefficiency of CLS lies in its failure to use the moment conditions (4.9c) and from its failure to use all of the moment conditions in (4.2c). The latter failure did not arise in the Ahn, Lee and Schmidt (2001) analysis (see footnote 2).

In Appendix D, we calculate the asymptotic variance matrix of the CLS estimator, under the "conditional independence of the moments up to fourth order" (CIM4) assumption of Section 4. Alternatively, along the lines of Appendix B, we could calculate our estimate of the asymptotic variance matrix that does not depend on the validity of the CIM4 assumption.

6 Empirical Application

This section includes an application of the estimators suggested in previous sections to the measurement of cost efficiency. The application uses panel data from Spanish savings banks covering the period 1992-1998. In order to allow for changes in cost efficiency over time, the individual effects are modeled in a parametric form as the "inverse" of the exponential time-varying function proposed by Battese and Coelli (1992) in a MLE framework.

[2]The moment conditions (5.5c) are equivalent to $E\Lambda' G(G'G)^{-1} b_{3i} = 0$. When the number of parameters in θ is less than $T-1$, the transformation $\Lambda' G(G'G)^{-1}$ loses information. This will be so in most parametric models for $\lambda(\theta)$, though it is not true in the model of Ahn, Lee and Schmidt (2001).

6.1 The Cost Frontier Model

The technology of banks is modeled using the following translog cost function:

$$
\begin{aligned}
\ln C_{it} &= \ln C^*(q_{it}, w_{it}, \gamma, \beta) + \alpha_{it} + \epsilon_{it} \\
&= \Bigg(\gamma + \sum_{j=1}^{m} \beta_{q_j} \ln q_{jit} + \frac{1}{2}\sum_{j=1}^{m} \beta_{q_j q_j} (\ln q_{jit})^2 + \sum_{j<k} \beta_{q_j q_k} \ln q_{jit} \ln q_{kit} \\
&\quad + \sum_{j=1}^{n} \beta_{w_j} \ln w_{jit} + \frac{1}{2}\sum_{j=1}^{n} \beta_{w_j w_j} (\ln w_{jit})^2 + \sum_{j<k} \beta_{w_j w_k} \ln w_{jit} \ln w_{kit} \\
&\quad + \sum_{j=1}^{m}\sum_{k=1}^{n} \beta_{q_j w_k} \ln q_{jit} \ln w_{kit}\Bigg) + \exp\big(\theta(t-1)\big)\alpha_i + \epsilon_{it}
\end{aligned}
\tag{6.1}
$$

where C_{it} is observed total cost, q_{jit} is the jth output, w_{kit} is the kth input price, β is a vector of parameters to be estimated, γ is a scalar to be estimated, and ϵ_{it} is the error term. The individual effects are modeled as the product of an exponential time-varying function $\lambda_t(\theta) = \exp\big(\theta(t-1)\big)$ and a time-invariant firm effect.

Our GMM estimators effectively treat the α_i as fixed. We can define the time-varying individual effects (intercepts) as $\alpha_{it} = \alpha_i \lambda_t(\theta)$. We wish to decompose the time-varying intercepts into a frontier intercept which varies over time (α_t) and a non-negative inefficiency term (v_{it}). That is:

$$\alpha_{it} = \lambda_t(\theta)\alpha_i = \alpha_t + v_{it},\ v_{it} \geq 0. \tag{6.2}$$

Following Cornwell, Schmidt and Sickles (1990) the frontier intercept can be estimated as:

$$\hat{\alpha}_t = \min_i(\hat{\alpha}_{it}) = \lambda_t(\hat{\theta}) \cdot \min_i(\hat{\alpha}_i) \tag{6.3}$$

and the inefficiency term as:

$$\hat{v}_{it} = \lambda_t(\hat{\theta})[\hat{\alpha}_i - \min_i(\hat{\alpha}_i)] = \lambda_t(\hat{\theta})\hat{v}_i \tag{6.4}$$

Since the dependent variable is expressed in natural logs, cost efficiency indexes can be calculated from (6.4) as:

$$\widehat{CE}_{it} = \exp\big(-\lambda_t(\hat{\theta})[\hat{\alpha}_i - \min_i(\hat{\alpha}_i)]\big) \tag{6.5}$$

It is easy to see from expressions (6.4) and (6.5) that cost efficiency compares the performance (individual effect) of a particular firm with a firm located on the frontier (i.e. with the minimum effect). Since cost efficiency is a relative concept, the average efficiency index is related to the variance σ_α^2: the higher the variance of the individual effects, the smaller the average efficiency.

6.2 Data

The application uses yearly data from Spanish savings banks. The number of banks has decreased over the last ten or fifteen years due to mergers and acquisitions. These mergers took place mainly in the early 1990s. In order to work with a balanced panel, we use data from 50 savings banks over the period 1992–98. Thus $N = 50$ and $T = 7$.

In Spain there are private banks as well as savings banks. We did not include private banks in our sample because private banks and savings banks are rather different. Savings banks concentrate on retail banking, providing checking, savings accounts and loan service to individuals (especially mortgage loans), whereas private banks are more involved in commercial and industrial loans. Another difference is that savings banks are more specialized than private banks in long-term loans, which do not require continuous monitoring. Since the two groups are likely to have different cost structures and have been regulated in different ways, we did not want to pool them. We tried to analyze private banks separately, but we were not very successful, and we do not report those results here.

The variables used in the analysis are defined as follows. We follow the majority of the literature and use the intermediation approach (proposed by Sealey and Lindley, 1977) which treats deposits as inputs and loans as outputs. We include three types of outputs and three types of inputs. The outputs are: Loans to banks, and other profitable assets (q_1); Loans to firms and households (q_2); and noninterest income (q_3). Using noninterest income goes beyond the intermediation approach as commonly modeled. We include it in an attempt to capture off-balance-sheet activities such as securitization, brokerage services, and management of financial assets for individual customers and mutual funds. This method of measuring nontraditional banking activities is not fully satisfactory. For example, we cannot distinguish between variation due to changes in volume and variation due to changes in price, and noninterest income is partly generated from traditional activities such as service charges on deposits or credits rather than from nontraditional activities.

The inputs are: Borrowed money, including demand, time and saving deposits, deposits from non-banks, securities sold under agreements to repurchase, and other borrowed money (x_1); Labor, measured by total number of employees (x_2); and Physical Capital, measured by the value of fixed assets in the balance sheet (x_3). All of the input prices, w_i ($i = 1, 2, 3$) were calculated in a straightforward way by dividing nominal expenses by input quantities. Accordingly, total cost includes both interest and operating expenses.

We normalize our variables by dividing by the sample geometric mean, so that the first order coefficients can be interpreted as the elasticities evaluated at that point. Furthermore, the standard homogeneity of degree one in input prices is imposed by normalizing cost and input prices using the price of physical capital as a numeraire. Thus in the end we have a translog function of five variables—three outputs and two normalized input prices. We also have an overall intercept, and our function $\lambda_t(\theta)$ has θ as a scalar. So, in terms of our previous notation, we have $N = 50$, $T = 7$, $K = 20$, $g = 1$ and $p = 1$. We have a total of 22 parameters (20 in β, 1 in γ, 1 in θ), in addition to the individual effects α_i ($i = 1, \ldots, 50$).

We note that the intrument vector W_i defined in Section 2 is of dimension $TK + g = 141$, which is rather large. For some of our methods of estimation we will instead make use of a reduced instrument set P_i, of dimension 6, which contains (for bank i) an intercept, plus the mean (over time) values of the three outputs and the two normalized input prices.

6.3 Methods of Estimation

The cost equation is estimated using the GMM estimators of the previous sections. We will use the following notation to refer to our estimators.

GMM1(W) is the GMM estimator based on the moment conditions (3.2). This estimator uses the Orthogonality assumption only. The **W** in the parentheses indicates that it uses the full instrument set W_i. There are 846 moment conditions. Since $N = 50 < 846$, $\hat{V}_{11}$ as given in equation (3.8) is singular—it is of rank 50 but dimension 846. In this case, and in other similar cases below, we used the Moore-Penrose pseudo-inverse $\hat{V}_{11}^{+}$ as our weighting matrix. For a justification of this procedure, see Doran and Schmidt (2002).

GMM1(P) is the GMM estimator that exploits the same Orthogonality assumption, but uses only the reduced instrument set P_i, that is, the moment conditions (A.1). There are 36 moment conditions so the estimated weighting matrix is non-singular (which is the point of using P_i).

GMM2 refers to the estimator based on the moment conditions (3.13), which were motivated by the NCH assumption. However, we evaluate the asymptotic variances without assuming NCH—see Appendix B. This is an exactly identified problem (number of moment conditions = number of parameters = 22) so the weighting matrix is irrelevant. We need estimates of the nuisance parameters in (3.13c), however. **GMM2(W)** will refer to the case that the nuisance parameter estimates are based on the GMM1(W) results, whereas **GMM2(P)** means that the GMM1(P) results were used.

GMM3(W) is the GMM estimator that is based on the moment conditions (4.2). This estimator relies on the Orthogonality and Covariance assumptions. The number of moment conditions is $846 + 20 + 6 = 872$. Once again we need a generalized inverse.

GMM3(P) uses the reduced instrument set P_i in (4.2a), and also omits the conditions (4.2b), in order to avoid having more instruments than banks. The number of moment conditions is $36 + 6 =$

42.

GMM4 is the estimator based on the moment conditions (4.9), (4.2b) and (4.2c). This set of moment conditions was motivated by the CIM4 assumption. The number of moment conditions is $22 + 20 + 6 = 48$. (We evaluate nuisance parameters in (4.9) using the GMM3(W) results so we could call our estimator **GMM4(W)**.) We evaluate the asymptotic variance without imposing the CIM4 assumption, as in Appendix B.

GMM5 is the estimator based on the moment conditions (4.9) and (4.2c). That is, compared to GMM4 we just drop the b_{2i} condition (4.2b). The number of moment conditions is 28. We will call this estimator **GMM5(W)** when the nuisance parameters are evaluated using the GMM3(W) results, and **GMM5(P)** when the nuisance parameters are evaluated using the GMM3(P) results.

CLS is the concentrated least squares estimator of Section 5. Its asymptotic variance is evaluated as in Appendix D.

In addition, we also consider the following estimators from the existing literature.

WITHIN is the standard within (fixed-effects) estimator that assumes time-invariant efficiency. See Schmidt and Sickles (1984).

MLE1 is a Battese-Coelli (1992)-type estimator. The model is still (6.1) but now we assume that ϵ_{it} is i.i.d. $N(0, \sigma_\epsilon^2)$ whereas α_i is i.i.d. as the truncation of $N(0, \sigma_\alpha^2)$. See Battese and Coelli (1992) for details of estimation of the inefficiencies.

MLE2 is an extension of MLE1, as follows. The model for MLE1 is not really comparable to our model for GMM, because the model for MLE1 has a time-invariant frontier. For MLE2 we assert that $\alpha_i = \alpha + u_i$, where u_i is truncated normal, but $\alpha \neq 0$. Now the error is $\epsilon_{it} + u_i \exp(\theta(t-1))$, so that we have a model of Battese-Coelli type, but the regression function (frontier) contains $\alpha \exp(\theta(t-1))$ which is time varying. An empirically relevant obseration is that α and γ are only separately identified when $\theta \neq 0$.

Comparing estimators, WITHIN can be viewed as a special case of GMM1. GMM1 relies on less assumptions than GMM3. The MLE2 results rely on stronger assumptions. In particular MLE2 makes the "random effects" assumption that α_i is independent of the regressors. Moreover, the individual effects α_i in the MLE model are also restricted to be i.i.d. half-normal (i.e., to be positive) and ϵ_{it} to be i.i.d. normal. The other GMM estimators (GMM2, GMM4, GMM5, CLS) represent attempts at simplication of GMM1 or GMM3. As noted, MLE1 is really non-comparable to the other estimators.

6.4 Empirical Results

The parameter estimates for the cost frontiers are presented in Table 1. All of the output coefficients are positive, except for one coefficient for GMM1(P), indicating that the estimated cost frontiers are, at the sample geometric mean, increasing in outputs. The input price elasticities are also positive at the geometric mean. These results confirm the positive monotonicity of the cost frontiers with respect to input prices and assure positive input cost-shares. Borrowed Money represents about 65 percent of banks' total cost, while Labor represents about 30 percent, and Physical Capital less than 5 percent.

Returns to scale can be estimated as one minus the scale elasticity; that is, as one minus the sum of the output cost elasticities. At the sample mean, the scale elasticity and returns to scale are only a function of the first-order output parameters. The results indicate moderately increasing returns to scale. However, this differs somewhat over methods of estimation. We can distinguish three groups of estimators in terms of their estimated returns to scale. For GMM2(P or W), GMM4, GMM5(P or W), CLS and MLE(1 or 2), this value is positive but less than 0.10, indicating the existence of moderate scale economies, as found in many past analyses of Spanish banks. These scale economies, however, are larger for WITHIN, GMM1(W) and GMM3(W), and they are unreasonably large for GMM1(P) and GMM3(P).

Next we consider the levels of cost efficiency implied by these estimates. The cost efficiency indexes are obtained using equation (6.5), except for the MLE estimators, where we follow Battese and Coelli (1992), with minor modifications to accomodate the case of a cost frontier instead of a production frontier.

Table 2 reports the yearly average of the estimated efficiency levels. For GMM4, GMM5(P or W), CLS and MLE(1 or 2) we obtain a mean efficiency index of approximately 85 or 90 percent. In these cases the average efficiency levels are similar to those found in other studies using the same data set (see, for instance, Cuesta and Orea, 2002). However, the scores are slightly lower for GMM1(W), GMM2(P or W) and GMM3(W), even lower for WITHIN, and ridiculously low for GMM1(P) and GMM3(P).

Except for the WITHIN estimator (where time-invariance of efficiency is imposed), the cost efficiency indexes increase or decrease over time on the basis of the sign of θ. If this parameter is positive (negative), efficiency decreases (increases) and the differences among firms increase (decrease) due to the exponential functional form of $\lambda_t(\theta)$. The estimated θ value is positive in all cases, and it is statistically different from zero except for GMM2(P or W), GMM1(P), GMM3(P) and MLE1. For GMM2(P or W), the estimated θ is not small, but its standard error is very large. Conversely, GMM1(W) and GMM3(W) give relatively small values of θ but they are still significant because their standard errors are small. Except for MLE1, the estimators that have the smallest values of θ (namely WITHIN, GMM1(P or W) and GMM3(P or W)) are also the ones that have the lowest efficiency scores.

In Table 3 we show the Spearman rank correlation coefficients between the efficiency levels estimated using the alternative models. Here the "observations" are the efficiency rankings of the various banks at a given time period (t). The efficiency levels vary over t but the ranks do not, so these rank correlations are the same for $t = 1, 2, \ldots, T$. The rank correlation coefficients are often high (in general, over 0.8), indicating that choosing a specific estimator is not necessarily a crucial issue when ranking firms in terms of their efficiency levels. However, WITHIN, GMM1(P) and

GMM3(P) are outliers, in the sense that their efficiencies do not correlate well with the efficiencies from the other methods, though they correlate strongly with each other. Recall that these three methods gave the lowest average efficiencies and also those that varied least over time, so they are outliers in many senses.

The estimated efficiency scores for the first period are graphed according to bank size in Figure 1. This figure seems to show a negative correlation between efficiency and size. We can see that again WITHIN, GMM1(P) and GMM3(P) are outliers, in that the efficiency levels are lower and they decrease more markedly with bank size. This is true to a lesser extent for GMM1(W) and GMM3(W). Even for the other methods, however, the efficiency levels are low for the very largest banks. This is especially apparent for the five biggest savings banks, which are the result of large mergers that took place in the early 1990's. Hence, these results support the argument that merged banks are less efficient than non-merged banks, as found also by Cuesta and Orea (2002).

Having discussed the economic results implied by our estimates, we now turn to a more econometrically oriented discussion. The basic issue is how to choose among these different estimators. We want to ask whether any of the models appear to be adequate and which seem to be best, in some sense. These are important and difficult questions in many efficiency estimation exercises, of course.

We will begin with the estimators that rely on the Orthogonality Assumption only, namely, GMM1 and GMM2. We can test the validity of the Orthogonality Assumption (and also the correctness of the regression specification) using the usual GMM overidentification test. A little notation will be useful here. Let N $(= 50)$ be the number of observations, s $(= 22)$ be the number of parameters, and r be the number of moment conditions (e.g., for GMM1(W), $r = 846$). A standard GMM result is that the minimized value of the GMM criterion function is asymptotically (for large N) distributed as chi-squared, with degrees of freedom equal to the degree of overidentificaion. This degree of overidentification is usually quoted as $r - s$, but this may not be appropriate when r

is larger than N. We will use $\min(r, N) - s$ as our degrees of freedom, since $\min(r, N)$ is the rank of the estimated variance matrix of the moment conditions, whose inverse (or generalized inverse) is used in calculating the criterion function. This distinction matters only when $N < r$, as is true for GMM1(W) and GMM3(W). Rejection of the joint hypothesis of correct specification and valid moment conditions is indicated by a significantly large chi-squared value.

For GMM2, the estimation problem is exactly identified so there is no overidentification test. For GMM1(W), we obtain a criterion function of 28.5, which is unremarkable in a chi-squared distribution with $50 - 22 = 28$ degrees of freedom. For GMM1(P), the statistic is 4.47, which is certainly not significantly large in a chi-squared distribution with $36 - 22 = 14$ degrees of freedom. In fact, the opposite is true: such a small value should arise with a probability of less than 0.01. This does not argue against the validity of the moment conditions but it indicates some problem, perhaps a failure of asymptotics to be reliable. We conclude that there is no strong evidence against the validity of the Orthogonality Assumption. However, having said that, we are not satisfied with the results that come from any of these estimators. The GMM1(P) results are simply not believable, as discussed above. For GMM1(W), the results are somewhat strange also, and we do not trust the accuracy of the asymptotic results, given the very large number of moment conditions. For GMM2(P or W), many of the results seem plausible, but the level of precision is too low to be acceptable. For example, the standard error for the estimate of θ is over ten times larger for GMM2 than for any other method. This may argue against the validity of the NCH assumption. However, it is also reasonable to wonder whether we might just have the "random effects" case that the effects are uncorrelated with all of the explanatory variables. In this case θ is not identified under the Orthogonality Assumption only.

Now consider estimators that use the Covariance Assumption in addition to the Orthogonality Assumption. We can test the hypothesis of the validity of the Covariance Assumption (as well as the model specification and the Orthogonality Assumption) using the GMM overidentification test for any of the GMM estimators that use these assumptions. GMM3(W) has a chi-squared statistic of 27.7, which is unremarkable in a chi-squared distribution with 28 degrees of freedom. GMM3(P) has a statistic of 10.2, which is suspiciously small, not large, with 20 degrees of freedom. GMM4 has a statistic of 39.4, which is significant at about the 0.05 level. GMM5(W) has a statistic of 7.98, which is not significant at usual levels. GMM5(P) has a statistic of 12.0, which is significant at about the 0.10 level. Thus there is some evidence, but not overwhelmingly strong evidence, against the Covariance Assumption.

We can also test the b_{2i} moment conditions (4.2b) separately from the b_{3i} moment conditions (4.2c). Here we use the general GMM result that, if we move from s moment conditions to $q > s$ conditions (and where the smaller set is a subset of the larger set), the difference in the GMM criterion functions is asymptotically chi-squared with $q - s$ degrees of freedom. To test the b_{3i} conditions, we compare GMM3(P) to GMM1(P), giving a statistic of $10.22 - 4.47 = 5.75$, which is not significant in a chi-squared with $42 - 36 = 6$ degrees of freedom. We can test the b_{2i} conditions by comparing GMM4 to GMM5(W), in which case the statistic is $39.4 - 8.0 = 31.4$, which is significant at about the 0.05 level in the chi-squared distribution with $28 - 8 = 20$ degrees of freedom. Again there is some evidence but not extremely strong evidence against the validity of the moment conditions implied by the Covariance Assumption.

Among these estimators, we would discard GMM3(P) on the basis of its too-small overidentification test statistic and its economically unbelievable results, and we would discard GMM3(W) because we do not trust the validity of the asymptotic results with so many moment conditions. Among the other estimators (GMM4, GMM5(P or W), and also CLS) there is less basis for choice, but fortunately it is also the case that the economically interesting results do not depend very much on which of these estimators one would choose. It is probably fair to say that GMM4 looks better than GMM5(P or W), since the estimates are quite similar but the level of precision is much better for GMM4. This argues against the assumption of normality of the errors (ϵ_{it}) since under the normality assumption GMM5 should be just as efficient as GMM4.

Finally, we turn to the MLE estimators. These depend on the Orthogonality and Covariance assumptions, but also on the "random effects" assumption and on distributional assumptions for the ϵ_{it} and the α_i. We can decisively reject the MLE1 model in favor of the MLE2 model, since the parameter α (see the discussion of MLE2 in Section 6.3) is very significantly different from zero. The MLE2 results are economically not terribly different from those from GMM4 (or CLS or GMM5, for that matter), a reassuring result in the sense that those models that we have trouble choosing between do not give strikingly different answers.

7 Conclusions

In this paper we have considered a panel data model with parametrically time-varying coefficients on the individual effects. Following Ahn, Lee and Schmidt (2001), we have enumerated the moment conditions implied by alternative sets of assumptions on the model. We have shown that our sets of moment conditions capture all of the useful information contained in our assumptions, so that the corresponding GMM estimators exploit these assumptions efficiently.

We have also considered concentrated least squares estimation. Here the incidental parameters problem is relevant because we are treating the fixed effects as parameters to be estimated. An interesting result is that the consistency of the least squares estimator requires both exogeneity assumptions and the assumption that the errors are white noise. Furthermore, given the white noise assumption, the least squares estimator is inefficient, because it fails to exploit all of the moment conditions that are available.

We show how the GMM estimation problem can be simplified under some additional assumptions, including the assumption of no conditional heteroskedasticity and a stronger conditional moment independence assumption. Under these assumptions we also give explicit expressions for the variance matrices of the GMM and least squares estimators.

Finally, we apply these GMM estimation methods to the measurement of cost efficiency of Spanish savings banks over the period 1992–1998. Different estimators give different results, and some estimators give very poor results. However, some of these estimators can be rejected on statistical grounds, and others give results that are too ridiculous to be taken seriously. Those methods that are not rejected on statistical or economic grounds give results that are similar to each other, and also to the results from previous analyses of these data. We conclude that our GMM methods can be practically useful.

APPENDIX

A GMM Using a Subset of the Instruments

First we consider estimation under the orthogonality assumption only. The available moment conditions are b_{1i} as given in equation (3.2). They assert that W_i in uncorrelated with $(u_{it} - \lambda_t u_{i1})$ for $t = 2, \ldots, T$, where W_i is the "instrument" vector that contains $X_{i1}, \ldots, X_{iT}$ and Z_i.

Now we suppose that we use only a subset of the instruments. So, let P_i be any subset of W_i, and we consider the moment conditions

$$Eb_{1i}^P(\beta, \gamma, \theta) = E[G(\theta)' u_i(\beta, \gamma) \otimes P_i] = 0, \tag{A.1}$$

which is the same as (3.2) except P_i replaces W_i. Now let $V_{11}^P = Eb_{1i}^P b_{1i}^{P\prime}$, of which a consistent estimate would be

$$\hat{V}_{11}^P = \frac{1}{N} \sum_{i=1}^{N} b_{1i}^P(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta}) b_{1i}^P(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta})' \tag{A.2}$$

where $\tilde{\beta}$, $\tilde{\gamma}$, $\tilde{\theta}$ is any consistent estimate. Let B_1^P be the matrix of expected derivatives of b_{1i}^P with respect to β, γ and θ. Then a consistent estimate of B_1^P is

$$\hat{B}_1^P = \frac{1}{N} \sum_{i=1}^{N} \hat{B}_{1i}^P, \quad \hat{B}_{1i}^P = -(\tilde{G}' X_i, \tilde{G}' 1_T Z_i', \tilde{u}_{i1} \tilde{\Lambda}_*) \otimes P_i \tag{A.3}$$

where $\tilde{G}$, $\tilde{u}_{i1}$ and $\tilde{\Lambda}_*$ are evaluated at any consistent estimate. Finally, $(\hat{B}_1^P \hat{V}_{11}^{P-1} \hat{B}_1^P)^{-1}$ is a consistent estimate of the asymptotic variance of $\sqrt{N}$ times $[(\hat{\beta} - \beta)', (\hat{\gamma} - \gamma)', (\hat{\theta} - \theta)']'$, where these are the GMM estimates from (A.1).

Next we consider estimation under the orthonality and covariance assumptions. In this case the set of available moment conditions consists of b_{1i}, b_{2i} and b_{3i} as given in equation (4.2). Now suppose that we use P_i instead of W_i, so that the moment conditions we use are expressed by b_{1i}^P, b_{2i} and b_{3i}. Let b_i^P be the vector that contains b_{1i}^P, b_{2i} and b_{3i}; let V^P be its variance matrix, and B^P

be its expected derivative matrix. We can estimate V^P as in (A.2) above, but with b^P_{1i} replaced by b^P_i. The expected derivative matrix B^P has blocks B^P_1, B_2 and B_3 (arranged vertically). B^P_1 can be estimated as in (A.3). B_2 and B_3 do not depend on whether we use P_i or W_i. We can estimate B_2 by

$$\hat{B}_2 = \frac{1}{N}\sum_{i=1}^{N}\hat{B}_{2i}, \text{ where} \tag{A.4a}$$

$$\hat{B}_{2i} = -H'[(\tilde{G}'X_i, \tilde{G}'1_T Z_i', \tilde{u}_{i1}\tilde{\Lambda}_*) \otimes \tilde{u}_i + \tilde{G}'\tilde{u}_i \otimes (X_i, 1_T Z_i', 0)]. \tag{A.4b}$$

Similarly we can estimate B_3 by

$$\hat{B}_3 = \frac{1}{N}\sum_{i=1}^{N}\hat{B}_{3i}, \text{ where} \tag{A.5a}$$

$$\hat{B}_{3i} = -\frac{\tilde{\lambda}'\tilde{u}_i}{\tilde{\lambda}'\tilde{\lambda}}(\tilde{G}'X_i, \tilde{G}'1_T Z_i', \tilde{u}_{i1}\tilde{\Lambda}_*) - \frac{1}{\tilde{\lambda}'\tilde{\lambda}}\tilde{G}'\tilde{u}_i\left(\tilde{\lambda}'X_i, \tilde{\lambda}'1_T Z_i', C_i\right) \tag{A.5b}$$

$$C_i = -\tilde{u}_i'\tilde{\Lambda} + 2\frac{\tilde{\lambda}'\tilde{u}_i}{\tilde{\lambda}'\tilde{\lambda}}\tilde{\lambda}'\tilde{\Lambda} \tag{A.5c}$$

B Variance Expressions That Do Not Depend on NCH or CIM4

First we will show how to calculate an estimate of the asymptotic variance matrix of the GMM estimate based on (3.13). The consistency of this estimate does not depend on the NCH assumption.

Our estimate will be of the form $(\hat{B}'\hat{V}^{-1}\hat{B})^{-1}$. Here $\hat{V}$ is an estimate of the variance matrix of the moment conditions, and it can be constructed in the usual way as the average over observations of the sums of squares and cross-products of the moment conditions. (That is, the calculation is analogous to (A.2) above, but with a different set of moment conditions.) Similarly, the estimate $\hat{B}$ is an avarage over observations of matrices of derivatives of the moment conditions with respect to the parameters. (That is, it is analogous to (A.3), (A.4) or (A.5) above.) To calculate these derivatives, we can simply replace the term $G'u_i$ in (3.13a), (3.13b) and (3.13c) by the matrix

$$-[\tilde{G}'X_i, \tilde{G}'1_T Z_i', \tilde{u}_{i1}\tilde{\Lambda}_*]. \tag{B.1}$$

In doing so, we are ignoring the fact that the terms preceeding $G'u_i$ depend on θ. But the derivatives of these terms would be multiplied by $G'u_i$, and the expectation of this product is zero, so they can be ignored.

Similar considerations apply to the GMM estimation based on (4.9), (4.2b) and (4.2c). We can construct an estimate of its variance matrix that does not depend on the correctness of the CIM4 assumption. The only complicated issue is the estimation of the expected derivative matrices. This follows from the previous discussion, as follows. (i) For the expected derivative matrix of (4.9), replace $G'u_i$ by $-(\tilde{G}'X_i, \tilde{G}'1_T Z_i', \tilde{u}_{i1}\tilde{\Lambda}_*)$ and average over observations. (ii) For the expected derivative matrix of (4.2b) and (4.2c), see (A.4) and (A.5) above.

C The Asymptotic Variance of the GMM Estimator Based on Orthogonality and Covariance Assumptions

Under the Orthogonality and Covariance Assumptions, the moment conditions we have are $b_{1i} = G'u_i \otimes W_i$, $b_{2i} = H'(G'u_i \otimes u_i)$, and $b_{3i} = (\lambda'\lambda)^{-1}\lambda' u_i \otimes G'u_i$. In Appendix A we showed how to estimate the variance of the GMM estimate based on these moment conditions. In this Appendix, we give an explicit expression of this variance matrix. The expression is calculated under the CIM4 assumption.

Let $\delta = (\beta', \gamma', \theta')'$. Let $B_j = -E(\partial b_{ji}/\partial\delta)$ for $j = 1, 2, 3$, evaluated at the true parameters. Let $V_{jk} = E b_{ji} b_{ki}'$ for $j, k = 1, 2, 3$, evaluated at the true parameters. Define $\kappa_3 = E\epsilon_{it}^3/\sigma_\epsilon^2$ and $\kappa_4 = E(\epsilon_{it}^4 - 3\sigma_\epsilon^4)/\sigma_\epsilon^2$. Let $\mu_W = EW_i$; $\Phi = \Phi(\theta) = \lambda_*\lambda_*' + \mathrm{diag}(\lambda_2, \ldots, \lambda_T)$; and $\Phi_* =$

$\lambda_*\lambda_*' + \mathrm{diag}(\lambda_2^2, \ldots, \lambda_T^2)$, where $\lambda_* = (\lambda_2, \ldots, \lambda_T)'$. After some algebra, we get

$$V_{11} = \sigma_\epsilon^2 (G'G \otimes \Sigma_{WW}) \tag{C.1}$$

$$V_{12} = \sigma_\epsilon^2 (G'G \otimes \Sigma_{W\alpha}\lambda')H \tag{C.2}$$

$$V_{13} = \sigma_\epsilon^2 \big[G'G \otimes \Sigma_{W\alpha} + \frac{\kappa_3}{\lambda'\lambda}(\Phi \otimes \mu_W)\big] \tag{C.3}$$

$$V_{22} = \sigma_\epsilon^2 H'[G'G \otimes (\sigma_\alpha^2 \lambda\lambda' + \sigma_\epsilon^2 I_T)]H \tag{C.4}$$

$$V_{23} = \sigma_\epsilon^2 H' \left\{ \left[\left(\sigma_\alpha^2 + \frac{\sigma_\epsilon^2}{\lambda'\lambda} \right) G'G + \frac{\kappa_3}{\lambda'\lambda} \mu_\alpha \Phi \right] \otimes \lambda \right\} \tag{C.5}$$

$$V_{33} = \sigma_\epsilon^2 \left\{ \left(\sigma_\alpha^2 + \frac{\sigma_\epsilon^2}{\lambda'\lambda} \right) G'G + 2\frac{\kappa_3}{\lambda'\lambda}\mu_\alpha \Phi + \frac{\kappa_4}{(\lambda'\lambda)^2}\Phi_* \right\} \tag{C.6}$$

and

$$B_1 = -[(G \otimes \Sigma_{WW})'S_X,\ (G \otimes \Sigma_{WW})'S_Z,\ \Lambda_* \otimes \Sigma_{W\alpha}] \tag{C.7}$$

$$B_2 = -H'(I_{T-1} \otimes \lambda)[(G \otimes \Sigma_{W\alpha})'S_X,\ (G \otimes \Sigma_{W\alpha})'S_Z,\ \sigma_\alpha^2 \Lambda_*] \tag{C.8}$$

$$B_3 = -[(G \otimes \Sigma_{W\alpha})'S_X,\ (G \otimes \Sigma_{W\alpha})'S_Z,\ \sigma_\alpha^2 \Lambda_*]. \tag{C.9}$$

With these results, the variance-covariance of the GMM estimator is

$$\mathrm{cov}\sqrt{N}(\hat{\delta} - \delta) = \left[(B_1',\ B_2',\ B_3') \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12}' & V_{22} & V_{23} \\ V_{13}' & V_{23}' & V_{33} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \right]^{-1}. \tag{C.10}$$

The evaluation of these results requires an estimate of σ_α^2, an issue not previously discussed. A very simple estimate of σ_α^2 can be obtained (assuming that the Covariance assumption holds) by solving (4.8) of the text. This yields

$$\hat{\sigma}_\alpha^2 = \frac{1}{\hat{\lambda}_t} \frac{1}{N} \sum_{i=1}^{N} \hat{u}_{it} \hat{u}_{i1} \tag{C.11}$$

for any single t (e.g. $t = 2$). A slightly more complicated estimator is as follows. Under the Covariance assumption, $G'\Sigma_{\epsilon\epsilon}G = \sigma_\epsilon^2 G'G$. Therefore, we can estimate σ_ϵ^2 by

$$\hat{\sigma}_\epsilon^2 = \mathrm{trace}(\widehat{G'\Sigma_{\epsilon\epsilon}G})/\mathrm{trace}(\hat{G}'\hat{G}) \tag{C.12}$$

where $\widehat{G'\Sigma_{\epsilon\epsilon}G}$ is given in equation (3.15), and $\mathrm{trace}(\hat{G}'\hat{G}) = \sum_{t=2}^{T}(1+\hat{\lambda}_t^2)$. Next, define $\hat{\Sigma}$ as the left hand side of equation (3.14); it is a consistent estimate of $\sigma_\epsilon^2 I_T + \sigma_\alpha^2 \lambda\lambda'$. So then a consistent estimate of σ_α^2 is

$$\tilde{\sigma}_\alpha^2 = \mathrm{trace}(\hat{\Sigma} - \hat{\sigma}_\epsilon^2 I_t)/\mathrm{trace}(\hat{\lambda}\hat{\lambda}') = [\mathrm{trace}(\hat{\Sigma}) - T\hat{\sigma}_\epsilon^2]/\hat{\lambda}'\hat{\lambda}. \tag{C.13}$$

D The asymptotic variance of the CLS estimator

In this Appendix we evaluate the asymptotic variance of the CLS estimator. We obtain an explicit expression by imposing the CIM4 assumption.

By the standard Taylor series expansion technique, we find that the asymptotic variance will be equal to $A_0 B_0^{-1} A_0$ where

$$A_0 = E\frac{\partial^2 C_i}{\partial\delta\partial\delta'}, \text{ and } B_0 = E\frac{\partial C_i}{\partial\delta}\frac{\partial C_i}{\partial\delta'} \tag{D.1}$$

evaluated at the true parameter. Let us calculate each of them. Let $\Lambda = \partial\lambda(\theta_0)/\partial\theta' = (0_{p\times 1},\ \Lambda_*')'$. B_0 is the same as in Ahn, Lee and Schmidt (2001, p. 253). Let $\Psi = G(G'G)^{-1}\Phi\cdot(G'G)^{-1}G'$; $\Psi_* = G(G'G)^{-1}\Phi_*(G'G)^{-1}G'$; and $\mu_\alpha = E\alpha_i$. Then

$$E\frac{\partial C_i}{\partial\beta}\frac{\partial C_i}{\partial\beta'} = 4\sigma_\epsilon^2 S_X'(P_G\otimes\Sigma_{WW})S_X \tag{D.2}$$

$$E\frac{\partial C_i}{\partial\beta}\frac{\partial C_i}{\partial\gamma'} = 4\sigma_\epsilon^2 S_X'(P_G\otimes\Sigma_{WW})S_Z \tag{D.3}$$

$$E\frac{\partial C_i}{\partial\beta}\frac{\partial C_i}{\partial\theta'} = 4\sigma_\epsilon^2 S_X'\left[P_G\otimes\Sigma_{W\alpha} + \frac{\kappa_3}{\lambda'\lambda}(\Psi\otimes\mu_W)\right]\Lambda \tag{D.4}$$

$$E\frac{\partial C_i}{\partial\gamma}\frac{\partial C_i}{\partial\gamma'} = 4\sigma_\epsilon^2 S_Z'(P_G\otimes\Sigma_{WW})S_Z \tag{D.5}$$

$$E\frac{\partial C_i}{\partial\gamma}\frac{\partial C_i}{\partial\theta'} = 4\sigma_\epsilon^2 S_Z'\left[P_G\otimes\Sigma_{W\alpha} + \frac{\kappa_3}{\lambda'\lambda}(\Psi\otimes\mu_W)\right]\Lambda \tag{D.6}$$

$$E\frac{\partial C_i}{\partial\theta}\frac{\partial C_i}{\partial\theta'} = 4\sigma_\epsilon^2\Lambda'\left\{\left(\sigma_\alpha^2 + \frac{\sigma_\epsilon^2}{\lambda'\lambda}\right)P_G + 2\frac{\kappa_3}{\lambda'\lambda}\mu_\alpha\Psi + \frac{\kappa_4}{(\lambda'\lambda)^2}\Psi_*\right\}\Lambda. \tag{D.7}$$

A_0 is obtained from the following.

$$E\frac{\partial^2 C_i}{\partial\beta\partial\delta'} = 2[S_X'(P_G \otimes \Sigma_{WW})S_X,\ S_X'(P_G \otimes \Sigma_{WW})S_Z,\ S_X'(P_G \otimes \Sigma_{W\alpha})\Lambda] \tag{D.8}$$

$$E\frac{\partial^2 C_i}{\partial\gamma\partial\delta'} = 2[S_Z'(P_G \otimes \Sigma_{WW})S_X,\ S_Z'(P_G \otimes \Sigma_{WW})S_Z,\ S_Z'(P_G \otimes \Sigma_{W\alpha})\Lambda] \tag{D.9}$$

$$E\frac{\partial^2 C_i}{\partial\theta\partial\delta'} = 2[\Lambda'(P_G \otimes \Sigma_{W\alpha}')S_X,\ \Lambda'(P_G \otimes \Sigma_{W\alpha}')S_Z,\ \sigma_\alpha^2\Lambda' P_G\Lambda]. \tag{D.10}$$

References

Ahn, S. C., Y. H. Lee, and P. Schmidt (2001) 'GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects.' *Journal of Econometrics* 101, 219–255.

Ahn, S. C. and P. Schmidt (1995) 'A Separability Result for GMM Estimation, with Applications to GLS Prediction and Conditional Moment Tests.' *Econometric Reviews* 14(1), 19–34.

Battese, G. E., and T. J. Coelli (1992) 'Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farms in India.' *Journal of Productivity Analysis* 3, 153–170.

Chamberlain, G. C. (1980) 'Analysis of Covariance with Qualitative Data.' *Review of Economic Studies* 47, 225–238.

Chamberlain, G. C. (1992) 'Efficiency Bounds for Semiparametric Regression.' *Econometrica* 60, 567–596.

Cornwell, C., P. Schmidt, and R. Sickles (1990) 'Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels.' *Journal of Econometrics* 46(1/2), 185–200.

Cuesta, R. A., and L. Orea (2002) 'Mergers and Technical Efficiency in Spanish Savings Banks: A Stochastic Distance Function Approach.' *Journal of Banking and Finance* 26, 2231–2247.

Doran, H. E., and P. Schmidt (2002) 'GMM Estimators with Improved Finite Sample Properties Using Principle Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model.' unpublished manuscript.

Hansen, L. P. (1982) 'Large Sample Properties of Generalized Method of Moments Estimators.' *Econometrica* 50, 1029–1054.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988) 'Estimating Vector Autoregressions with Panel Data.' *Econometrica* 56, 1371–1395.

Kiefer, N. M. (1980) 'A Time Series–Cross Sectional Model with Fixed Effects with an Intertemporal Factor Structure.' unpublished manuscript, Cornell University.

Kumbhakar, S. C. (1990) 'Production Frontiers, Panel Data and Time-varying Technical Inefficiency.' *Journal of Econometrics* 46, 201–212.

Lee, Y. H. (1991) 'Panel Data Models with Multiplicative Individual and Time Effects: Applications to Compensation and Frontier Production Functions.' unpublished Ph.D. dissertation, Department of Economics, Michigan State University.

Lee, Y. H., and P. Schmidt (1993) 'A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency.' In *The Measurement of Productive Efficiency: Techniques and Applications,* edited by H. Fried, C. A. K. Lovell, and S. Schmidt, Oxford University Press.

Schmidt, P., and R. Sickles (1984) 'Production Frontiers and Panel Data.' *Journal of Business and Economic Statistics* 2(4), 367-374.

Sealey, C. W., and J. T. Lindley (1977) 'Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions.' *Journal of Finance* 32(4), 1251-1266.

Table 1. Estimated coefficients.

	WITHIN		GMM1(W)		GMM2(W)		GMM3(W)		GMM4(W)		GMM5(W)		CLS	
	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.
$\ln q_1$	0.284	22.60	0.314	12.28	0.323	10.65	0.302	11.88	0.320	56.97	0.328	33.67	0.309	44.21
$\ln q_2$	0.491	24.45	0.423	9.97	0.540	5.60	0.430	9.77	0.548	46.86	0.578	30.32	0.543	47.01
$\ln q_3$	0.036	2.74	0.120	4.39	0.062	3.44	0.109	4.33	0.062	6.49	0.063	4.24	0.057	7.00
$\ln w_1$	0.694	61.27	0.634	49.71	0.668	87.33	0.634	49.98	0.685	168.09	0.723	66.08	0.678	91.77
$\ln w_2$	0.292	21.10	0.328	14.05	0.286	18.68	0.332	14.57	0.299	37.54	0.285	20.30	0.305	35.58
$.5(\ln q_1)^2$	0.165	6.79	0.191	2.76	0.202	8.25	0.184	2.76	0.182	10.12	0.200	8.67	0.177	12.38
$.5(\ln q_2)^2$	0.136	2.29	0.029	0.22	0.138	2.18	0.016	0.12	0.106	3.70	0.120	2.11	0.102	2.98
$.5(\ln q_3)^2$	0.021	0.47	-0.009	-0.09	-0.077	-1.48	-0.055	-0.58	0.043	1.12	0.042	0.95	0.000	0.01
$.5(\ln w_1)^2$	0.108	4.48	0.140	3.82	0.136	1.11	0.134	3.73	0.179	10.78	0.160	5.48	0.141	9.63
$.5(\ln w_2)^2$	0.055	1.16	0.218	2.40	0.110	2.24	0.211	2.24	0.086	3.12	0.088	1.62	0.065	2.28
$\ln q_1 \ln q_2$	-0.119	-4.22	-0.121	-1.71	-0.202	-7.55	-0.132	-1.91	-0.108	-6.75	-0.118	-5.02	-0.126	-7.61
$\ln q_1 \ln q_3$	-0.039	-1.72	-0.061	-1.66	0.005	0.12	-0.044	-1.32	-0.082	-6.69	-0.095	-5.13	-0.054	-4.22
$\ln q_1 \ln w_1$	0.075	4.36	0.017	0.39	0.047	0.95	0.023	0.59	0.037	3.05	0.008	0.36	0.050	4.65
$\ln q_1 \ln w_2$	-0.062	-2.44	0.028	0.44	-0.014	-0.13	0.021	0.40	0.029	2.11	0.036	1.37	-0.004	-0.24
$\ln q_2 \ln q_3$	-0.003	-0.06	0.043	0.48	0.049	0.83	0.070	0.77	0.008	0.23	0.019	0.41	0.017	0.63
$\ln q_2 \ln w_1$	-0.130	-5.08	-0.035	-0.80	-0.035	-0.28	-0.041	-1.05	-0.057	-3.15	-0.009	-0.31	-0.086	-5.23
$\ln q_2 \ln w_2$	0.068	1.77	-0.120	-2.01	-0.022	-0.32	-0.122	-2.20	-0.014	-0.53	-0.026	-0.65	0.036	1.52
$\ln q_3 \ln w_1$	0.060	2.63	0.016	0.36	0.003	0.07	0.019	0.46	0.040	2.23	0.008	0.28	0.046	3.24
$\ln q_3 \ln w_2$	0.005	0.15	0.065	1.14	0.024	0.55	0.070	1.31	-0.004	-0.15	0.032	0.79	-0.013	-0.63
$\ln w_1 \ln w_2$	-0.064	-2.45	-0.125	-2.64	-0.133	-0.98	-0.120	-2.62	-0.156	-11.36	-0.168	-5.75	-0.116	-6.96
Intercept	-	-	9.968	132.45	9.811	23.68	9.962	155.64	9.745	1539.55	9.688	885.52	9.814	391.43
θ	-	-	0.045	2.81	0.082	0.32	0.042	3.08	0.097	23.05	0.160	15.38	0.063	9.34
α	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RTS	0.188		0.143		0.075		0.159		0.070		0.031		0.091	
σ_α			0.298				0.298		0.104		0.063		0.159	
σ_ϵ			0.040				0.040		0.017		0.022		0.015	
Obj. Function			28.526				27.694		39.412		7.98		0.0032	
d.f.			28				28		26		6			

Table 1. Estimated coefficients(cont).

	MLE1		MLE2		GMM1(P)		GMM2(P)		GMM3(P)		GMM5(P)	
	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.
$\ln q_1$	0.316	28.65	0.304	31.50	0.156	2.85	0.322	23.73	0.164	4.75	0.322	36.58
$\ln q_2$	0.592	27.56	0.590	34.75	0.480	7.62	0.575	11.38	0.454	8.52	0.548	27.21
$\ln q_3$	0.032	2.37	0.047	3.95	-0.016	-0.24	0.053	2.78	0.071	1.61	0.054	3.59
$\ln w_1$	0.754	73.10	0.683	60.26	0.642	19.45	0.675	45.11	0.642	24.04	0.681	68.77
$\ln w_2$	0.245	18.22	0.305	24.23	0.283	3.53	0.292	7.72	0.249	3.51	0.303	21.76
$.5(\ln q_1)^2$	0.181	7.42	0.176	8.25	-0.135	-1.09	0.174	5.34	-0.173	-1.78	0.174	8.14
$.5(\ln q_2)^2$	0.122	1.70	0.068	1.41	0.445	2.33	0.175	1.39	0.336	1.96	0.103	1.78
$.5(\ln q_3)^2$	0.052	0.86	0.014	0.37	0.161	1.17	-0.002	-0.03	0.152	1.08	0.018	0.35
$.5(\ln w_1)^2$	0.140	5.78	0.143	6.45	0.066	1.22	0.140	1.09	-0.003	-0.05	0.149	7.35
$.5(\ln w_2)^2$	0.053	1.12	0.028	0.68	0.159	1.27	0.062	0.93	0.335	2.13	0.055	1.12
$\ln q_1 \ln q_2$	-0.101	-3.41	-0.104	-4.08	-0.018	-0.10	-0.175	-3.50	0.029	0.21	-0.117	-5.26
$\ln q_1 \ln q_3$	-0.079	-3.41	-0.076	-3.87	0.191	2.72	-0.006	-0.07	0.153	2.30	-0.058	-3.09
$\ln q_1 \ln w_1$	0.061	3.54	0.051	3.13	0.238	3.19	0.044	0.51	0.274	4.46	0.040	2.42
$\ln q_1 \ln w_2$	-0.046	-1.77	-0.013	-0.60	-0.135	-2.00	0.000	0.00	-0.092	-1.52	0.020	0.94
$\ln q_2 \ln q_3$	-0.004	-0.06	0.037	1.04	-0.400	-2.76	-0.008	-0.08	-0.334	-2.44	0.007	0.13
$\ln q_2 \ln w_1$	-0.142	-5.01	-0.100	-3.95	-0.238	-2.59	-0.052	-0.32	-0.228	-2.84	-0.069	-2.65
$\ln q_2 \ln w_2$	0.082	2.19	0.066	1.99	-0.095	-1.01	0.029	0.34	-0.221	-2.06	0.001	0.03
$\ln q_3 \ln w_1$	0.082	3.13	0.063	2.72	0.011	0.22	0.022	0.56	-0.016	-0.37	0.046	1.90
$\ln q_3 \ln w_2$	-0.016	-0.44	-0.027	-0.87	0.221	2.29	-0.010	-0.23	0.255	2.85	-0.008	-0.22
$\ln w_1 \ln w_2$	-0.103	-3.86	-0.117	-4.81	-0.032	-0.49	-0.121	-0.78	-0.002	-0.03	-0.121	-5.96
Intercept	9.504	648.74	9.807	291.45	7.751	0.44	9.750	44.05	10.112	128.04	9.758	520.37
θ	0.001	0.05	0.075	6.38	0.002	0.12	0.126	0.46	0.005	1.03	0.087	8.78
α	-	-	-0.230	-6.77	-	-	-	-	-	-	-	-
RTS	0.060		0.059		0.380		0.050		0.311		0.076	
σ_α	0.140		0.112		1.893		0.095		0.607		0.109	
σ_ϵ	0.027		0.024		0.045		0.020		0.052		0.018	
Obj. Function	668.06		706.55		4.472				10.221		11.953	
d.f.					14				20		6	

Table 2. Average efficiency (in percentage).

YEAR	WITHIN	GMM1 (W)	GMM2 (W)	GMM3 (W)	GMM4 (W)	GMM5 (W)	CLS	MLE1	MLE2	GMM1 (P)	GMM2 (P)	GMM3 (P)	GMM5 (P)
$t = 1$	57.25	82.73	86.33	82.70	89.89	91.54	88.37	89.00	91.03	27.58	89.27	36.64	89.13
$t = 2$	57.25	82.05	85.28	82.07	88.93	90.16	87.68	89.00	90.38	27.50	87.93	36.47	88.22
$t = 3$	57.25	81.35	84.15	81.42	87.90	88.58	86.96	88.99	89.69	27.43	86.44	36.29	87.24
$t = 4$	57.25	80.63	82.95	80.75	86.78	86.77	86.19	88.98	88.95	27.35	84.79	36.12	86.18
$t = 5$	57.25	79.89	81.67	80.06	85.56	84.70	85.39	88.98	88.17	27.27	82.96	35.95	85.05
$t = 6$	57.25	79.12	80.31	79.35	84.25	82.36	84.55	88.97	87.33	27.19	80.94	35.78	83.84
$t = 7$	57.25	78.33	78.87	78.63	82.84	79.70	83.67	88.97	86.45	27.12	78.71	35.61	82.55
$t = 1, \ldots, 7$	57.25	80.59	82.79	80.71	86.59	86.26	86.12	88.98	88.86	27.35	84.43	36.12	86.03

Table 3. Spearman rank correlation coefficients (in percentage).

	WITHIN	GMM1 (W)	GMM2 (W)	GMM3 (W)	GMM4 (W)	GMM5 (W)	CLS	MLE1	MLE2	GMM1 (P)	GMM2 (P)	GMM3 (P)	GMM5 (P)
WITHIN	100												
GMM1(W)	86.3	100											
GMM2(W)	74.9	93.0	100										
GMM3(W)	91.2	98.9	90.6	100									
GMM4(W)	72.8	90.9	98.2	88.2	100								
GMM5(W)	36.1	66.6	83.6	59.9	87.0	100							
CLS	80.5	95.3	98.4	93.8	98.2	80.0	100						
MLE1	75.2	84.1	92.8	83.8	94.9	81.7	94.4	100					
MLE2	69.1	86.5	96.9	84.4	98.4	88.1	96.4	96.0	100				
GMM1(P)	95.6	78.5	63.6	84.5	59.7	20.5	69.3	61.2	56.1	100			
GMM2(P)	57.2	82.0	95.5	77.6	96.4	93.2	92.3	90.4	97.1	44.6	100		
GMM3(P)	95.9	82.3	68.8	87.4	64.6	27.0	73.8	66.5	61.5	98.6	51.2	100	
GMM5(P)	73.5	91.2	98.4	88.8	99.8	86.1	98.5	94.7	98.3	61.0	96.1	65.8	100

Figure 1. Efficiency levels at first period.